ISBORD: Internet Searching based on Resource Description☆

Zhiwei Yan[a], Sherali Zeadally[b], Guanggang Geng[a,*]

[a] *China Internet Network Information Center, NANEL, Beijing, 100190, PR China*
[b] *College of Communication and Information, University of Kentucky, Lexington, KY, 40506, USA*

Abstract

Based on the Information-Centric Networking (ICN) concept and the mature model of the current TCP/IP-based Internet, we propose content searching based on universal and scalable resource description, namely ISBORD (Internet Searching based on Resource Description). This novel concept aims to improve the efficiency of content searching and simplifies the end-user functionality to support the evolution of the content-centric Internet.

1. Introduction

The current Internet architecture cannot effectively and efficiently meet various requirements, including security, mobility, scalability, and quality of service because of several shortcomings in its original design [1]. A key goal of the original Internet design was to efficiently interconnect mainframes and minicomputers, and to provide efficient remote access to them. This end-to-end approach and especially its specific practical implementation, however, have been identified as the root causes of several key limitations of the current Internet architecture. Then various add-ons, such as Network Address Translation (NAT), Mobile IP (MIP), Content Distribution Network (CDN), Peer-to-Peer (P2P) overlay, etc., all violate, in various ways, several aspects of the original Internet architecture in order to provide features that were not part of its original requirements.

Moreover, it has been observed recently that information is at the heart of almost all communications and particularly Internet uses [2]. This information-centric usage of the Internet raises various architectural challenges, many of which are not effectively handled by the current Internet architecture, even with various add-ons mentioned above. One direct and unified approach to address this problem is to replace the "where is the information?" with "what is the information?". Then the Information-Centric Networking (ICN) concept was proposed to address the question of "what is the information?" [3].

In the past few years, various ICN projects [4–7] have proposed ICN architectures that leverage in-network storage for caching, multiparty communication through replication, and interaction models that decouple senders and receivers. The common goal of these proposals is to achieve efficient and reliable distribution of content by providing a general platform for communication services that are today only available in dedicated systems such as P2P overlay and proprietary CDN. The ICN approach is also expected to be able to better cope with disconnections, disruptions, and flash crowd effects while supporting various types of communication services [8,9].

A naming system is also used in ICN to efficiently manage all the contents as in the TCP/IP-based Internet. However, in this naming system, it is difficult for the user to identify the name of the content directly and to invoke content addressing, although name-based routing is efficient and secure.

In contrast, the Domain Name System (DNS) which was created for mapping names to addresses of hosts and networks, and keeps evolving with the explosive extension of the global Internet. Today, there are more than 300 million second level
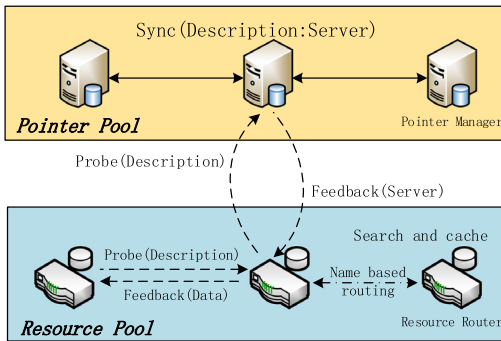
Fig. 1. ISBORD communication model.

Fig. 2. Structure of the resource description.

domain names covering more than 1000 top level domains. DNS offers more functionalities than the original "host table" because there could be multiple records associated with a given fully-qualified name. This flexibility and scalability demonstrate how the DNS can act as the basic and core naming system in the current Internet.

But it is also impossible for current Internet users to remember so many DNS names to retrieve the related information. Fortunately, search engine is more user-friendly to map the DNS name with the key words which can be input by Internet users as the information profile, without being aware of the DNS names that are actually present. It is conceivable that search engine use becomes so ubiquitous that ready access to DNS-based navigation mechanisms will disappear entirely from user interfaces. But as an application-layer solution, search engine is another add-on, which poses additional security, scalability and efficiency challenges to the current Internet.

Based on the above considerations, in this work we propose a novel Internet architecture, Internet Searching based on Resource Description (ISBORD), to support the direct fuzzy search of content based on the description, which is quickly becoming an emerging Internet trend.

2. ISBORD

For scalability and efficiency, ISBORD consists of two layers: the management layer and the routing layer. To differentiate ISBORD from the current Internet architecture design, the communication model consists of two layers namely, the *Pointer Pool* and the *Resource Pool* as shown in Fig. 1.

Pointer pool: This layer should be involved with the management of the registration information of Internet resources. When a publisher wants to push some content to the Internet, the related information such as the <Description: Server> pair will be registered with a Pointer Manager. "Description" is the keyword to identify this resource, and "Server" is the server name of the content provider. The Pointer Managers will synchronize this information as the CDN nodes do for example.

Resource pool: The Resource Routers in this layer are in charge of content addressing and caching based on the request from the users. Normally, two types of routing should be supported: description-based routing and name-based routing. In the former case, the router checks its local cache first and responds
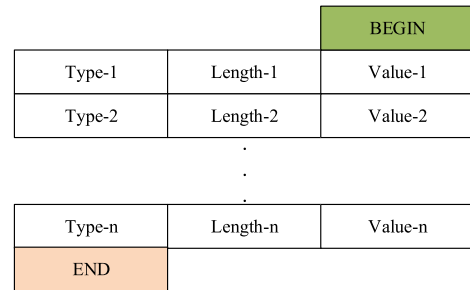
back with the matched data as today's search engine does. If the request cannot be satisfied within the requested hops, the Pointer Manager requests the content directly from the server. In the latter case, the content name is used to fetch the content directly from the specified server as the router in Named Data Networking (NDN) [5] does. The difference between ISBORD and NDN is that description-based routing is used in ISBORD to improve the user experience and avoid the name resolution overlay in NDN.

2.1. Signaling messages

To support this communication model, the following signaling messages should be used:

Probe: This message is used to request the content based on the description contained inside it. Besides, a Time-To-Live (TTL) should be set in the Probe message to specify the hop limit of this message. When the Resource Router receives the Probe message, it checks its local cache with the description information. If there is no positive match, the Resource Router propagates the Probe message with the TTL decreased by one. When the value of TTL is zero, the Resource Router should send the Probe message to its default Pointer Manager.

Feedback: This message is the response to the Probe message. For the Resource Router, the related Probe message carries the content that matches with the description. For the Pointer Manager, the name of the content server is contained in order to enable the Resource Router to fetch the requested content directly from the server.

Sync: This message is multicast or broadcast based and it is used to synchronize the content description and the related server's name among the Pointer Managers.

2.2. Resource description

The way we describe the resource in a scalable and universal manner is crucial for the large-scale deployment of the ISBORD. The structure of the resource description is shown in Fig. 2.

Starting with BEGIN and ending with END, part of the description consists of one or more keywords. The keyword, being the basic element, is organized by the Type-Length-Value (TLV) field. For example, if a user wants to fetch a picture of